Apolar surface area determines the efficiency of translocon-mediated membrane-protein integration into the endoplasmic reticulum

Karin Öjemalm[a], Takashi Higuchi[b], Yang Jiang[c], Ülo Langel[c], IngMarie Nilsson[a], Stephen H. White[d], Hiroaki Suga[b,1], and Gunnar von Heijne[a,e,1]

[a]Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden; [b]Department of Chemistry, Graduate School of Science, The University of Tokyo, 113-0033 Tokyo, Japan; [c]Department of Neurochemistry, Stockholm University, SE-10691 Stockholm, Sweden; [d]Department of Physiology and Biophysics and the Center for Biomembrane Systems, University of California, Irvine, CA 92697-4560; and [e]Science for Life Laboratory Stockholm University, SE-171 77 Solna, Sweden

Edited by Donald Engelman, Yale University, New Haven, CT, and approved April 12, 2011 (received for review January 6, 2011)

Integral membrane proteins are integrated cotranslationally into the membrane of the endoplasmic reticulum in a process mediated by the Sec61 translocon. Transmembrane α-helices in a translocating polypeptide chain gain access to the surrounding membrane through a lateral gate in the wall of the translocon channel [van den Berg B, et al. (2004) *Nature* 427:36–44; Zimmer J, et al. (2008) *Nature* 455:936–943; Egea PF, Stroud RM (2010) *Proc Natl Acad Sci USA* 107:17182–17187]. To clarify the nature of the membrane-integration process, we have measured the insertion efficiency into the endoplasmic reticulum membrane of model hydrophobic segments containing nonproteinogenic aliphatic and aromatic amino acids. We find that an amino acid's contribution to the apparent free energy of membrane-insertion is directly proportional to the nonpolar accessible surface area of its side chain, as expected for thermodynamic partitioning between aqueous and nonpolar phases. But unlike bulk-phase partitioning, characterized by a nonpolar solvation parameter of 23 cal/(mol · Å^2), the solvation parameter for transfer from translocon to bilayer is 6–10 cal/(mol · Å^2), pointing to important differences between translocon-guided partitioning and simple water-to-membrane partitioning. Our results provide compelling evidence for a thermodynamic partitioning model and insights into the physical properties of the translocon.

flexizyme | hydrophobicity | nonproteinogenic amino acid | transmembrane helix

In eukaryotic cells, membrane proteins destined for the plasma membrane and the various compartments along the endo- and exocytic pathways are synthesized by endoplasmic reticulum (ER)-bound ribosomes and cotranslationally integrated into the ER membrane in a process mediated by the Sec61 translocon complex; the homologous SecYEG translocon mediates membrane-protein integration into the inner membrane of prokaryotes (1, 2). Subsequent to membrane integration, membrane proteins fold and oligomerize in the ER and are then moved further along the secretory pathway by vesicular transport.

During the membrane-integration step, hydrophobic segments in the translocating nascent polypeptide chain exit the Sec61 translocon through a lateral gate and become embedded in the surrounding lipid bilayer (3–5). Cotranslational, translocon-mediated integration of transmembrane α-helices into the ER membrane sets the stage for all subsequent folding and oligomerization events and hence represents a critical step in the maturation of membrane proteins.

In previous studies, we have provided quantitative data on the propensities of the 20 natural amino acids to promote the integration of transmembrane helices into the ER membrane and have shown that they depend both on hydrophobicity and on position within the helix (3, 6). Although the partitioning of transmembrane helices between the Sec61 translocon and the lipid membrane bears strong similarities to partitioning of solutes between water and lipid membranes, translocon-to-bilayer partitioning may not be equivalent to water-to-bilayer partitioning (7). Insights into the differences between the two partitioning processes might be revealed if the physicochemical properties of the translocon could be probed chemically. However, given the somewhat idiosyncratic collection of proteinogenic amino acids used in nature, it has hitherto not been possible to vary side-chain chemistry in the systematic fashion required to unravel fully the physicochemical basis for translocon-mediated membrane partitioning. In order to probe the membrane-integration mechanism in greater detail, we have taken advantage of a suppressor tRNA-based technique to introduce nonproteinogenic aliphatic and aromatic amino acids into a model hydrophobic segment and measure their apparent free energies of membrane insertion. The results show not only that translocon/membrane partitioning is quantitatively different from simple water/membrane partitioning, but also reveal a physicochemical asymmetry between the cytoplasmic and lumenal ends of a transmembrane helix, which is likely related to the structure of the translocon.

Results

Experimental Approach. The basic approach is illustrated in Fig. 1. A suppressor tRNA ($tRNA_{sup}$) is charged with the desired nonproteinogenic amino acid using the Flexizyme system (8, 9) and is then added to an in vitro translation system programmed with an mRNA encoding an engineered version of the well-characterized membrane-protein leader peptidase (Lep). The Lep construct has two N-terminal transmembrane helices (TM1, TM2) and contains a hydrophobic test segment (H segment) flanked by two canonical Asn-X-Thr acceptors sites for N-linked glycosylation (G1, G2). A UAG stop codon that can be recognized by the charged $tRNA_{sup}$ serves to position the nonproteinogenic amino acid in the H segment.

When the translation reaction is carried out in the presence of ER-derived dog pancreas rough microsomes (RMs), TM1 and TM2 insert into the ER membrane as shown (10). If the H segment is recognized as a transmembrane segment by the translocon and inserted into the membrane, only the G1 acceptor site is

Author contributions: K.Ö., T.H., H.S., and G.v.H. designed research; K.Ö. and T.H. performed research; Y.J., U.L., and I.M.N. contributed new reagents/analytic tools; K.Ö., T.H., S.H.W., and G.v.H. analyzed data; and K.Ö., T.H., S.H.W., H.S., and G.v.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. E-mail: gunnar@dbb.su.se or hsuga@chem.s.u-tokyo.ac.jp.

See Author Summary on page 12571.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1100120108/-/DCSupplemental.

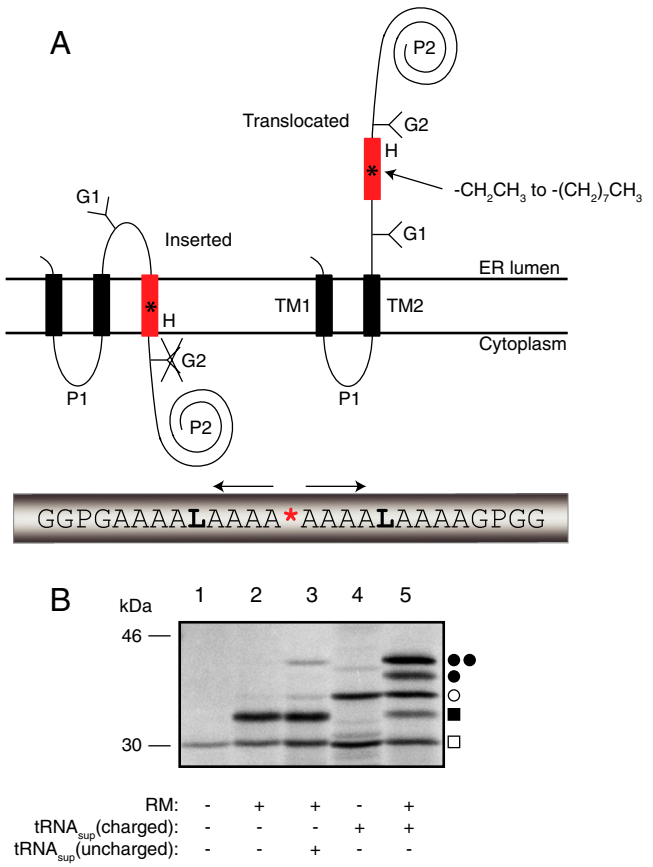

Fig. 1. (*A*) Model H segments are introduced into the Lep host protein between two engineered acceptor sites for N-linked glycosylation (G1, G2). A typical H segment is shown, including a nonproteinogenic amino acid (*) and GGPG…GPGG flanking regions. The fractions of membrane-inserted and noninserted H segment are determined by quantitation of radiolabeled singly and doubly glycosylated forms of the protein. (*B*) In vitro translation of a Lep construct with the H segment AAAALAAAAAAAAALAAAX (X = Me-Trp) in the absence and presence of charged or uncharged $tRNA_{sup}$ and dog pancreas RMs. □, Truncated Lep protein resulting from termination at the UAG stop codon in the H segment; ■, truncated Lep protein glycosylated on the G1 site; ○, full-length, nonglycosylated Lep protein; ●, full-length Lep protein glycosylated on the G1 site; ●●, full-length Lep protein glycosylated on the G1 and G2 sites. The degree of suppression in the absence of added $tRNA_{sup}$ is negligible (lane 2), and is very low (4%) with added uncharged $tRNA_{sup}$ (lane 3). Suppression is >50% in the presence of charged $tRNA_{sup}$ (lanes 4 and 5). The amounts of sample loaded in −RM and +RM lanes were adjusted to give roughly equal signals.

accessible to the lumenal oligosaccharyl transferase enzyme and receives a glycan moiety; if, in contrast, the H segment is translocated across the membrane, both G1 and G2 become glycosylated (11). The apparent free energy of membrane insertion of a given H segment is calculated as $\Delta G_{app} = -RT \ln(f_1/f_2)$, where R is the gas constant, T the absolute temperature (298 K), f_1 the amount of singly glycosylated molecules, and f_2 the amount of doubly glycosylated molecules (6). The H segments analyzed here have the composition GGPG-[1X,nL,(18-n)A]-GPGG, where X is the nonproteinogenic amino acid and n is chosen such that, for any given X, -1 kcal/mol $\leq \Delta G_{app} \leq +1$ kcal/mol. The GGPG…GPGG flanks are included in order to break any regular secondary structure and thereby insulate the H segment from the influences of the surrounding sequence and to prevent it from shifting in position along the membrane normal (6).

Membrane-Insertion Characteristics of Nonpolar Side Chains. Because hydrophobicity is central to transmembrane-helix integration into the ER membrane, we first asked how nonpolar surface area correlates with propensity for membrane insertion. The most straightforward way to determine the relation between nonpolar surface area and ΔG_{app} is to choose amino acids with linear alkyl side chains for X. We therefore scanned amino acids with linear side chains containing two to eight carbons along an H segment with $n = 2$ leucines. The results are shown in Fig. 2*A* and listed in Table S1. Three important conclusions can be drawn: (*i*) ΔG_{app} varies in a regular fashion with the length of the side chain; (*ii*) the difference in ΔG_{app} between the terminal and central positions of the side chain increases with the length of the side chain; and (*iii*) for the longer side chains, there is a noticeable asymmetry in the curves when one compares positions close to the N-terminal end with those close to the C-terminal end of the H segment (especially evident comparing positions 6 and 14, Fig. 2 *A* and *D*). Cyclic aliphatic and nonpolar aromatic side chains yield similar results, Figs. 2*B* and 3*A*.

As seen in Fig. 2*C*, there is a strict correlation between the increase in side-chain accessible surface area (ASA) and ΔG_{app} that holds for aliphatic side chains with at least up to three times the ASA of the largest natural aliphatic amino acids. For the middle position in the H segment, ΔG_{app} decreases by approximately 10 cal/(mol · Å^2), whereas the decrease is smaller when the side-chain is closer to the ends of the H segment, Fig. 2*D*. A similar relation between ASA and ΔG_{app} holds for the nonpolar aromatic side chains, but the decrease in ΔG_{app} is only about 7 cal/(mol · Å^2) for the middle position of the H segment in this case, Fig. 2*C*.

Except for the slight asymmetry in the curves, these results are those expected if membrane insertion of the H segment is driven by thermodynamic partitioning into a lipid bilayer environment. First, similar to our results, the partitioning free energy of nonpolar compounds between aqueous buffer and a bilayer-mimetic solvent is proportional to ASA and can be characterized by the atomic solvation parameter σ. For partitioning between buffer and nonpolar solvents such as n-octanol, σ is approximately 23–25 cal/(mol · Å^2) for aliphatic side chains and approximately 16 cal/(mol · Å^2) for nonpolar aromatic side chains (12, 13). The corresponding values for translocon/bilayer partitioning in the middle of the membrane, approximately 10 and 7 cal/(mol · Å^2), are 2.5-fold smaller, but the value for nonpolar aromatic side chains is approximately 70% of the value for aliphatic side chains in both cases. Second, large aliphatic compounds are expected to partition preferentially into the center of the bilayer (14), as has been shown experimentally for n-hexane (15), and we see the same trend in our data, Fig. 2*D*.

Membrane-Insertion Characteristics of Polar Aromatic Side Chains. We also tested a selection of polar aromatic side chains, some containing non-carbon atoms, Fig. 3. As found previously, Tyr and Trp behave differently compared to non-polar aromatic residues such as Phe in that they promote membrane insertion significantly better when located near the ends of the H segment than when in the middle (3, 6); the same behavior is now seen for the aniline side chain but not for methylated tyrosine, methylated tryptophan, or the benzothiophene side chain. This observation points to the hydrogen-bonding ability of the Tyr, Trp, and aniline side chains (and the lack thereof in the Phe, methyl-Tyr, methyl-Trp, and benzothiophene side chains) as a critical distinguishing factor. A study of the interactions of Trp analogs with and without H-bonding ability showed that the "aromaticity" of Trp was the dominant cause of the preferential partitioning of Trp into lipid bilayer interfaces (16). Our results indicate that H bonding becomes important when polar aromatic residues are inserted in locations below the membrane-water interface.

Transmembrane Asymmetry. Finally, what could be the cause of the slight transbilayer asymmetry in the ΔG_{app} curves in Fig. 2*A*?

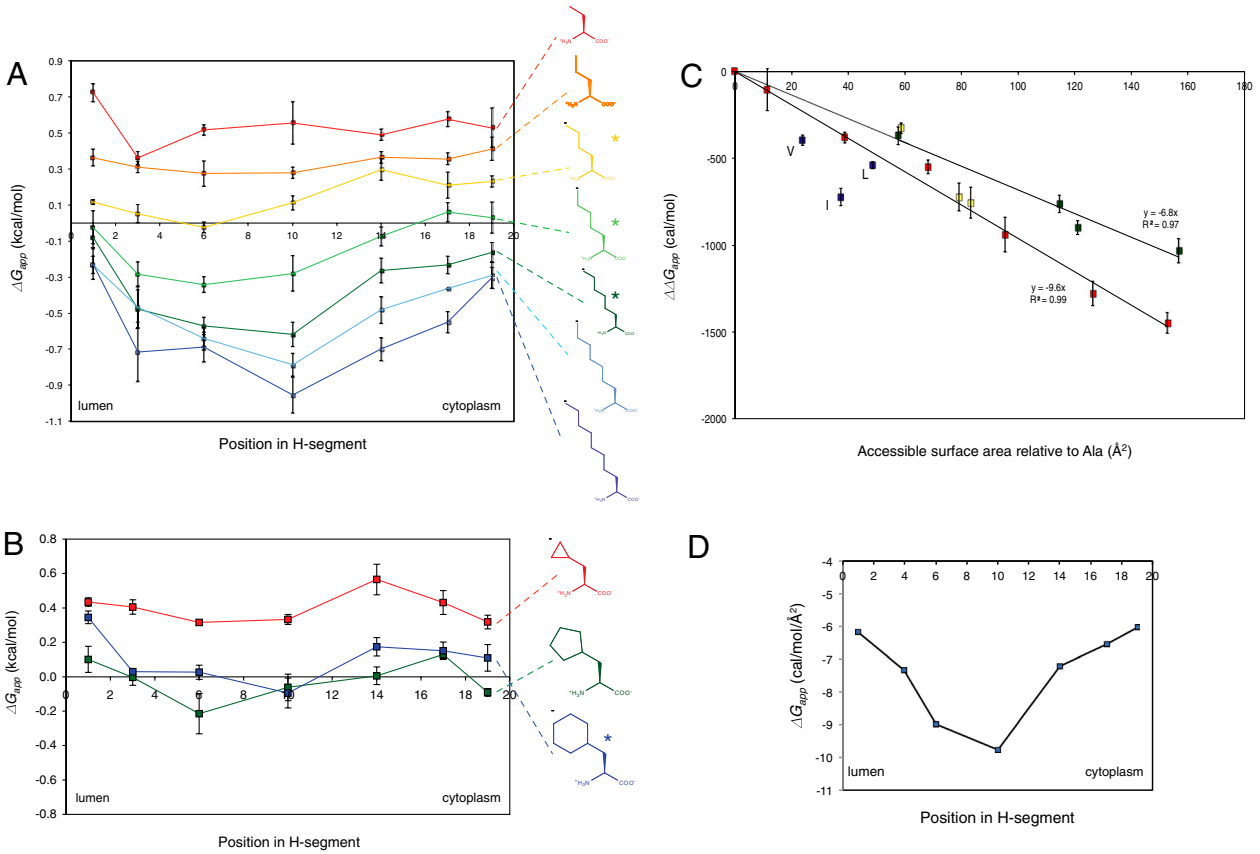

Fig. 2. (*A*) Apparent free energy of insertion (ΔG_{app}) for 19-residue-long H segments carrying a single nonproteinogenic amino acid with a linear alkyl side chain (shown on the right) in the indicated positions. The H segments all have the sequence AAAALAAAAAAAAALAAAA with the Ala in the indicated position replaced by the nonproteinogenic amino acid. The lumenal, N-terminal end of the H segment is on the left and the cytoplasmic end on the right. Error bars show standard deviations; * indicates side chains for which the ΔG_{app} value for position 6 is significantly smaller than the value for position 14 (two-sided *t* test, $p < 0.01$), i.e., that have a significant asymmetry in the ΔG_{app} profile. (*B*) ΔG_{app} for cyclic alkyl side chains. The H segments all have the sequence AAAALAAAAAAAAALAAAA with the Ala in the indicated position replaced by the nonproteinogenic amino acid. (*C*) ΔG_{app} decreases in proportion to the increase in ASA of the nonproteinogenic amino acid. $\Delta\Delta G_{app}$ is measured relative to the AAAALAAAAAAAAALAAAA H segment and the nonproteinogenic amino acid is in the middle position (position 10) in the H segment. The change in ASA is calculated relative to a model AAAALAAAAAAAAALAAAA α-helix. Red data points are for linear alkyl side chains, yellow for cyclic alkyl side chains (cyclopropyl, cyclopentyl, cyclohexyl), green for aromatic side chains lacking polar groups (Phe, 1-naphthyl, 2-naphthyl, biphenyl), and blue for the branched natural amino acids Leu, Ile, and Val. The average change in ΔG_{app} is –9.6 cal/(mol · Å^2) for the linear alkyl side chains and –6.8 cal/(mol · Å^2) for the nonpolar aromatic side chains. (*D*) Average change in ΔG_{app} per square angstrom (Å^2) of accessible surface area as a function of position in the H segment (calculated from the data in *A* using linear regression as in *C*).

We asked whether the asymmetry correlates with the N-to-C-terminal polarity of the H segment or with the orientation of the H segment relative to the membrane. To invert the membrane orientation of H segments containing hexyl and octyl side chains in different positions, we inserted an additional GGPG-[7L,12A]-GPGG transmembrane segment between TM2 and the H segment, Fig. 4. Comparing Figs. 4 and 2*A*, there is a general trend that the membrane-insertion efficiency is higher when the alkyl side chain is in the lumenal half of the H segment (i.e., the asymmetry in the ΔG_{app} curves correlates with the membrane orientation of the H segment), suggesting that it reflects an asymmetry in either the ER membrane or the Sec61 translocon. Although data are scarce, the ER is thought not to have a strong asymmetry in lipid composition between the two leaflets (17, 18). Therefore, it is more likely that the origin of the asymmetry is to be sought in the translocon. Although we cannot know the exact nature of the translocon asymmetry until a high-resolution structure of a functionally open translocon becomes available, the recent structure of a translocon with a partially open lateral gate (5) reveals the nascent-chain conduit as a deep canyon lined mainly by apolar residues and with scattered polar groups projecting from the canyon walls, Fig. 5. It does not seem unreasonable that the complex shape and physicochemical heterogeneity of the translocon channel might underlie the observed asymmetry in the ΔG_{app} curves.

Discussion

Our results provide compelling evidence for the thermodynamic partitioning model of translocon-mediated integration of transmembrane helices into the ER membrane (3, 6, 19–21). In its simplest version, this model pictures the transmembrane helix as equilibrating between the translocon channel and the surrounding membrane. As judged from the available X-ray structures of prokaryotic homologs of the Sec61 translocon (4, 5, 22, 23), the channel is quite narrow and lined by a mixture of polar and apolar amino acids, providing an environment that is less polar than aqueous buffer. Likewise, the ER membrane, which, like all biological membranes, has a high protein content, may offer an environment that is less apolar than a pure lipid bilayer (24). Indeed, the solvation parameters for partitioning of aliphatic and nonpolar aromatic side chains obtained here [approximately 10 and 7 cal/(mol · Å^2)] are a factor 2.5 smaller than found in classical solute transfer experiments (13, 25–28), suggesting that simple water-to-lipid partitioning measurements do not capture the full complexity of translocon-to-membrane partitioning (7). The asymmetry in the ΔG_{app} curves, Figs. 2 and 4, and effects on the hydrophobicity threshold for H segment insertion caused

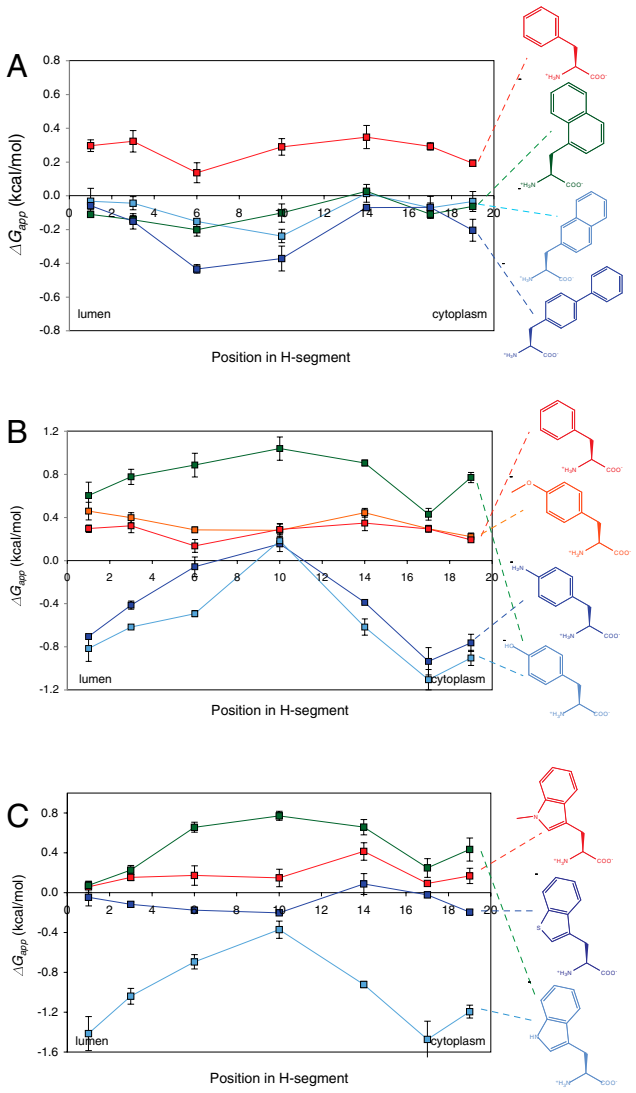

Fig. 3. ΔG_{app} for different aromatic amino acids. The H segments used in *A* have the sequence AAAALAAAAAAAAALAAAA, with the Ala in the indicated position replaced by the nonproteinogenic amino acid. In *B*, the H segments used for Phe, Me-Tyr, and Tyr (green trace) also have this sequence, whereas the sequence used for Tyr (light-blue trace) and the aniline side chain is AAAALALAAAAALALAAAA. In *C*, AAAALAAAAAAAAALAAAA is used for Trp (green trace), Me-Trp and the benzothiophene side chain, and AAAALALAAAAALALAAAA is used for Trp (light-blue trace).

by mutations in the Sec61 translocon (29) support the notion that our measurements report on partitioning between the translocon channel and the surrounding lipid, although we cannot completely rule out more complicated models where the H segment can also explore the membrane-water interface region in the vicinity of the translocon during the membrane-insertion step.

Little is known about the energetics of the passage of nascent membrane-protein chains through the Sec61 translocon. Is the chain pushed steadily through the translocon by the ribosome? Or does the chain diffuse through the translocon aided by accessory proteins, such as BiP, which give directionality via a Brownian ratchet mechanism (30, 31)? If the chain is driven steadily through the translocon by the ribosome, Schow et al. (7) have suggested that there are only two independent equilibria that need be considered: one between translocon and bilayer (ΔG_{tbi}) and one between water and bilayer (ΔG_{wbi}). The first equilibrium process determines whether a transmembrane helix enters the membrane and the second determines if it stays there. Schow et al. (7) further estimated that $\Delta G_{wbi} \approx 2.6 \Delta G_{tbi}$, and

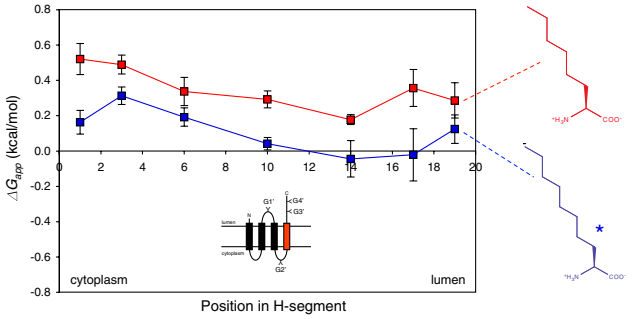

Fig. 4. ΔG_{app} for N_{cyt}-C_{lum} orientated linear hexyl and octyl side chains. The GGPG…GPGG flanked H segments are inserted C terminal to an added third transmembrane helix (GGPGALAALALAALAALALAALAGPGG) in the Lep host protein and acceptor sites for N-linked glycosylation are engineered into the construct, as shown in the cartoon. If the H segment does not insert into the membrane, only the G1′ site will receive a glycan; if it does insert, the G3′ and G4′ sites will also be modified. The H segments all have the composition [*X*,18A] and have the nonproteinogenic amino acid *X* in the indicated position. The cytoplasmic, N-terminal end of the H segment is on the left and the lumenal on the right. The ΔG_{app} value for position 3 is significantly larger than the value for position 17 (two-sided *t* test, $p < 0.01$) for the octyl side chain (indicated by *).

suggested that the discrepancy could be explained by assuming that the translocon/bilayer solvation parameter $\sigma_{tbi} = \sigma_{wbi}/2.6 = 8.8$ cal/(mol · Å^2). This value is remarkably close to the value for σ_{tbi} (approximately 10 cal/(mol · Å^2)) reported here. Although the agreement may be fortuitous, it does suggest a possible route to understanding quantitatively the physical chemistry of translocon-to-membrane partitioning.

Materials and Methods

Enzymes and Chemicals. All enzymes were purchased from Fermentas, except Phusion DNA polymerase from Finnzyme and SP6 RNA polymerase from Promega. The QuikChange™ Site-Directed Mutagenesis kit and deoxyribonucleotides were from Stratagene, and the Megashortscript™ T7 kit was from Ambion Inc. The plasmid pGEM1, the rabbit reticulocyte lysate system, and the RNasin were from Promega. Oligonucleotides were from Eurofins MWG Operon. All chemicals were from Sigma-Aldrich, except DMSO from J.T. Baker Chemicals, ethanol from Kemetyl, and [^{35}S]-methionine from PerkinElmer.

DNA Manipulations. For cloning, a modified version of the *Escherichia coli lepB* gene in a pGEM1 vector was used (6). It harbors an engineered test segment (H segment) in the coding region of the P2 domain between a Spe1 cleavage site in codons 226–227 and a Kpn1 cleavage site in codon 253 (WT *lepB* codon positions) as well as two glycosylation acceptor sites for N-linked glycosylation at codons 96–98 (G1: Asn-Ser-Thr) and codons 258–260 (G2: Asn-Ala-Thr). The sequences of the H segment's flanking regions are QETKENGIRLSETSGGPG-(H segment)-GPGGVPGQQNATWIVPP (Spe1 and Kpn1 cleavage sites underlined). The introduction of amber stop codons (TAG) into the termini of the H-segment encoding sequence was done by site-directed mutagenesis using Pfu Turbo polymerase. To introduce amber stop codons in or near the center of the H segment, double-stranded oligonucleotides encoding the H segment with the amber stop (including also GGPG/GPGG) and flanked by N-terminal Spe1 and C-terminal Kpn1 sticky ends were first generated by annealing of two pairs of complementary oligonucleotides with overlapping overhangs (each 18–45 nucleotides long), followed by annealing of the pairs via the complementary overhangs and cloning into the *lepB* gene between the Spe1 and Kpn1 cleavage sites (6).

For Lep constructs with a N_{cyt}-C_{lum} orientated H segment, a large part of the P2 domain was replaced by the corresponding part of the P2 domain of a Lep construct where all positive charges after codon 179 until 30 codons upstream of the end of the protein were replaced by alanines, and a new H segment with composition GGPG-ALAALALAALAALALAALA-GPGG was introduced between Apa1 and Mfe1 cleavage sites, located in codons −71 and −45 relative to the first codon of the original H segment. The construct further has two additional glycosylation acceptor sites for N-linked glycosylation, one (G2′: Asn-Ala-Thr) 16–18 codons downstream of the new H segment and the other (G4′: Asn-Ser-Thr) 7–9 codons downstream of G3′ (the G3′ site is the same as the G2 site in the original construct with

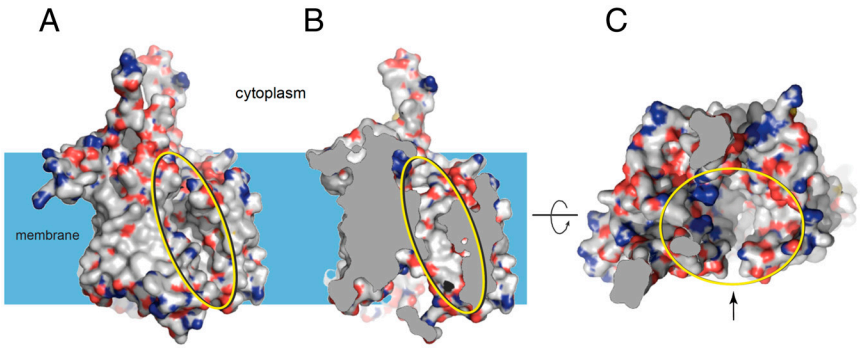

Fig. 5. Different views of the nascent-chain conduit in *Pyrococcus furiosus* SecYEβ (5) whose structure reveals a partially open ("primed") translocon. (*A*) View along the membrane plane, showing the partially open lateral gate (yellow oval). (*B*) Same view as in *A* but with the front of the molecule removed to show the rear wall of the channel. (*C*) View of the upper parts of the channel (yellow oval) seen from the cytoplasmic end. The lateral gate is indicated by the arrow. Carbon atoms are displayed in white, oxygen atoms in red, nitrogen atoms in blue, and sulfur atoms in yellow.

N_{out}-C_{in} orientated H segment). The gene fragment of the replacing P2 domain was generated by PCR amplification and inserted between a Xho1 cleavage site (*lepB* codons 180 and 181) and a Sma1 cleavage site directly downstream of the 3′ end of the *lepB* gene.

Preparation of Flexible tRNA Aminoacylation Ribozyme [Dinitrobenzyl Flexizyme (dFx); Enhanced Flexizyme (eFx)] and Microbacteriophage L5 (ML)-Derived $tRNA^{Asn}_{cta}$ [ML-$tRNA^{Asn}_{cta}$ ($tRNA_{sup}$)]. Preparations were done using the same protocol. First, double-stranded DNA templates encoding the RNA species and an N-terminal T7 promotor sequence (8, 32) were generated by PCR extension of annealed overlapping oligonuclotides. DNA templates were then amplified by PCR using primers complementary to both ends of the templates, followed by phenol/chloroform extraction and ethanol precipitation. The DNA was used in a second step for transcription by T7 polymerase using the Ambion Megashortscript™ T7 kit, and the RNA product was isopropanol precipitated and purified over 12% denaturing PAGE. After cutting out the RNA band, RNA was eluted for 2 h in 0.3 M NaCl, ethanol precipitated, and dissolved in dH_2O (70–250 μM final concentration).

Materials for the Synthesis and Characterization of Amino Acid Derivatives. All experiments dealing with air- and moisture-sensitive compounds were conducted under an atmosphere of dry argon. For TLC analysis, Merck precoated plates (silica gel 60 F_{254}, Art 5715, 0.25 mm) were used. For flash column chromatography, silica gel 60 (Merck Art 7734, 70–230 mesh) was used. For silica gel preparative TLC preparation, Merck precoated plates (silica gel 60 F_{254}, Art 5744, 0.5 mm) were used. Cyanomethylester (CME) and 3,5-dinitrobenzylester derivatives were prepared using a previously described procedure (8, 33). For general procedures of synthesis and characterization of compounds, see *SI Text*.

Acylation of ML-$tRNA^{Asn}_{cta}$. Flexizyme (dFx in all cases, except for aromatic amino acids where eFx was used) and ML-$tRNA^{Asn}_{cta}$ ($tRNA_{sup}$), each 250 μM in 71 mM Hepes-K buffer, pH 7.5 (total volume 7 μL), were heated at 95 °C for 2 min and cooled to room temperature over 5 min. One microliter $MgCl_2$ (200 mM for dFx and 3 M for eFx reactions) and 2 μL of amino acid 3,5-dinitrobenzyl ester (25 mM, nonaromatic amino acids) or amino acid CME (25 mM, aromatic amino acids) were added. The reaction was carried out for 2 h on ice and then stopped with 40 μM AcONa (0.3 M, pH 5.2). RNA was ethanol precipitated twice, once with a 0.1 M ethanolic AcONa solution (pH 5.2) and once with 95% ethanol. More detailed information can be found elsewhere (8).

Expression in Vitro and Quantification of Membrane-Insertion Efficiency. All constructs were transcribed for 60 min at 37 °C using a standard SP6 polymerase transcription protocol (34). Resulting mRNA was translated for 80 min at 30 °C in rabbit reticulocyte lysate (40–60 pg mRNA per μL translation mix) in the presence of [^{35}S]-Met (370 pCi/μL translation mix; 1 Ci = 37 GBq), dog pancreas RMs (50 pL/μL translation mix), amino acid mix (each 75 μM/μL translation mix), amino acid-$tRNA_{sup}$ (3.5 pmol/μL translation mix), and RNasin (3 units per μL translation mix). Translation products were analyzed by SDS-PAGE. Gels were visualized on a Fuji FLA-3000 PhosphoImager using the Image Reader 8.1j software and quantified using ImageGauge V 3.45 and the Qtiplot 0.9.3-rc2 softwares (35). The degree of membrane integration of each H segment was calculated as an apparent equilibrium constant between the membrane-integrated and nonintegrated forms: $K_{app} = f_1/f_2$, where f_1 is the fraction of singly and f_2 the fraction of doubly glycosylated Lep molecules, and the results were then converted to apparent free energies, $\Delta G_{app} = -RT \ln K_{app}$. All reported ΔG_{app} values are averages of three or four independent measurements. The degree of suppression of the UAG stop codon obtained in the in vitro translation system varied between 50% and 75% for the different aminoacylated $tRNA_{sup}$ species, and was between 2% and 6% (all positions shown in Fig. 2 were tested) when an uncharged $tRNA_{sup}$ was present in the translation mix.

Accessible Surface Area Calculations. Model helices of the H segments without flanking GGPG/GPGG sequences were generated using MacPyMol and, if needed, the side chains of the nonnatural amino acids manually modeled at the appropriate position in the helix. The accessible surface area was then calculated using Naccess version 2.1.1 (copyright S. Hubbard and J. Thornton 1992–1996) with a rolling probe size of 1.4 Å.

ACKNOWLEDGMENTS. We thank Drs. Kálmán Szabó and Erik Lindahl, Stockholm University, for technical advice, and Dr. Berhard Dobberstein, Heidelberg University, for kindly providing dog pancreas rough microsomes. This work was supported by grants from the European Research Council (ERC-2008-AdG 232648), the Swedish Foundation for Strategic Research, the Swedish Research Council, and the Swedish Cancer Foundation (G.v.H.), the Swedish Foundation for International Cooperation in Research and Higher Education and Henrik Granholms Stiftelse (I.M.N.), the Japan Science and Technology Agency, Strategic International Cooperative Program (H.S.), and the US National Institutes of Health (S.H.W.).

1. Rapoport TA (2007) Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* 450:663–669.
2. White SH, von Heijne G (2008) How translocons select transmembrane helices. *Annu Rev Biophys* 37:23–42.
3. Hessa T, et al. (2007) Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* 450:1026–1030.
4. van den Berg B, et al. (2004) X-ray structure of a protein-conducting channel. *Nature* 427:36–44.
5. Egea PF, Stroud RM (2010) Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. *Proc Natl Acad Sci USA* 107:17182–17187.
6. Hessa T, et al. (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* 433:377–381.
7. Schow EV, et al. (2011) Arginine in membranes: The connection between molecular dynamics simulations and translocon-mediated insertion experiments. *J Membr Biol* 239:35–48.
8. Murakami H, Ohta A, Ashigai H, Suga H (2006) A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nat Methods* 3:357–359.
9. Xiao H, Murakami H, Suga H, Ferre-D'Amare AR (2008) Structural basis of specific tRNA aminoacylation by a small in vitro selected ribozyme. *Nature* 454:358–361.
10. Johansson M, Nilsson I, von Heijne G (1993) Positively charged amino acids placed next to a signal sequence block protein translocation more efficiently in *Escherichia coli* than in mammalian microsomes. *Mol Gen Genet* 239:251–256.
11. Sääf A, Wallin E, von Heijne G (1998) Stop-transfer function of pseudo-random amino acid segments during translocation across prokaryotic and eukaryotic membranes. *Eur J Biochem* 251:821–829.
12. Wimley WC, Creamer TP, White SH (1996) Solvation energies of amino acid sidechains and backbone in a family of host-guest pentapeptides. *Biochemistry* 35:5109–5124.

13. Karplus PA (1997) Hydrophobicity regained. *Protein Sci* 6:1302–1307.
14. Marqusee JA, Dill KA (1986) Solute partitioning into chain molecule interphases: Monolayers, bilayer membranes, and micelles. *J Chem Phys* 85:434–444.
15. White SH, King GI, Cain JE (1981) Location of hexane in lipid bilayers determined by neutron diffraction. *Nature* 290:161–163.
16. Yau WM, Wimley WC, Gawrisch K, White SH (1998) The preference of tryptophan for membrane interfaces. *Biochemistry* 37:14713–14718.
17. Buton X, Morrot G, Fellmann P, Seigneuret M (1996) Ultrafast glycerophospholipid-selective transbilayer motion mediated by a protein in the endoplasmic reticulum membrane. *J Biol Chem* 271:6651–6657.
18. van Meer G, Voelker DR, Feigenson GW (2008) Membrane lipids: Where they are and how they behave. *Nat Rev Mol Cell Biol* 9:112–124.
19. Heinrich S, Mothes W, Brunner J, Rapoport T (2000) The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell* 102:233–244.
20. Gumbart J, Chipot C, Schulten K (2011) Free-energy cost for translocon-assisted insertion of membrane proteins. *Proc Natl Acad Sci USA* 108:3596–3601.
21. Rychkova A, Vicatos S, Warshel A (2010) On the energetics of translocon-assisted insertion of charged transmembrane helices into membranes. *Proc Natl Acad Sci USA* 107:17598–17603.
22. Zimmer J, Nam Y, Rapoport TA (2008) Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* 455:936–943.
23. Tsukazaki T, et al. (2008) Conformational transition of Sec machinery inferred from bacterial SecYE structures. *Nature* 455:988–991.
24. Johansson AC, Lindahl E (2009) The role of lipid composition for insertion and stabilization of amino acids in membranes. *J Chem Phys* 130:185101.
25. Chothia C (1976) The nature of the accessible and buried surfaces in proteins. *J Mol Biol* 105:1–12.
26. Eisenberg D, McLachlan AD (1986) Solvation energy in protein folding. *Nature* 319:199–203.
27. Hermann RB (1977) Use of solvent cavity area and number of packed solvent molecules around a solute in regard to hydrocarbon solubilities and hydrophobic interactions. *Proc Natl Acad Sci USA* 74:4144–4145.
28. Reynolds JA, Gilbert DB, Tanford C (1974) Empirical correlation between hydrophobic free energy and aqueous cavity surface area. *Proc Natl Acad Sci USA* 71:2925–2927.
29. Junne T, Kocik L, Spiess M (2010) The hydrophobic core of the Sec61 translocon defines the hydrophobicity threshold for membrane integration. *Mol Biol Cell* 21:1662–1670.
30. Simon SM, Peskin CS, Oster GF (1992) What drives the translocation of proteins? *Proc Natl Acad Sci USA* 89:3770–3774.
31. Liebermeister W, Rapoport TA, Heinrich R (2001) Ratcheting in post-translational protein translocation: A mathematical model. *J Mol Biol* 305:643–656.
32. Ohta A, Murakami H, Higashimura E, Suga H (2007) Synthesis of polyester by means of genetic code reprogramming. *Chem Biol* 14:1315–1322.
33. Saito H, Kourouklis D, Suga H (2001) An in vitro evolved precursor tRNA with aminoacylation activity. *EMBO J* 20:1797–1806.
34. Nilsson I, Johnson AE, von Heijne G (2003) How hydrophobic is alanine? *J Biol Chem* 278:29389–29393.
35. Hedin LE, et al. (2010) Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. *J Mol Biol* 396:221–229.

Supporting Information

Öjemalm et al. 10.1073/pnas.1100120108

SI Text

Compound Characterization. ^{1}H NMR and ^{13}C NMR were measured on a Bruker AV300 spectrometer. Chemical shifts are expressed in ppm downfield from internal standard (tetramethylsilane, 0.00 ppm), and coupling constants are reported as hertz. Splitting patterns are indicated as follows: br, broad; s, singlet; d doublet; t, triplet; q, quartet; m, multiplet. High-resolution mass spectra (HRMS) were obtained with Thermo Exactive spectrometer.

General Procedure for the Synthesis of Amino Acids with 3,5-Dinitrobenzyl Ester (*Procedure A*). *(S)-3,5-Dinitrobenzyl 2-Amino-3-Cyclopropylpropanoate Hydrochloride (for the Introduction of Three-Membered-Ring Side Chain).* The following is a representative experimental procedure of procedure A: To a solution of 2-amino-3-cyclopropylpropanoic acid (30 mg, 0.23 mmol) and Boc_2O (61 mg, 0.28 mmol) in dioxane (0.4 mL) and aqueous 1 M NaOH (0.4 mL), $NaHCO_3$ (48 mg, 0.57 mmol) was added at room temperature. After stirring for 5 h, the reaction mixture was diluted by adding ethyl ethanoate (EtOAc). The mixture was washed by saturated aqueous $KHSO_4$ and extracted with EtOAc (x3). The combined organic extracts were washed with brine, dried ($MgSO_4$), and concentrated in vacuo. To a solution of this crude mixture in dimethylformamide (DMF) (0.8 mL), 2, 4-dinitrobenzylchloride (65 mg, 0.30 mmol), and diisopropylethylamine (42 mg, 0.057 mL, 0.42 mmol) were added at room temperature. After stirring for 16 h, the reaction was quenched by adding saturated aqueous $KHSO_4$, and mixture was extracted with Et_2O (x2). The combined organic extracts were washed with water (x2), saturated $NaHCO_3$ (x2) and brine, dried ($MgSO_4$), and concentrated in vacuo. The residue was purified by preparative thin-layer chromatography (PTLC) (hexane/acetone = 4/1) to afford 3,5-dinitrobenzylester (DNB) derivative (40 mg, 44%). To a solution of DNB derivative (20 mg, 0.051 mmol) in EtOAc (0.5 mL), 4 M HCl in EtOAc (2 mL) was added at 0 °C. After stirring for 10 h, starting material was disappeared and the reaction mixture was concentrated in vacuo (16 mg, 95%). ^{1}H NMR (300 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 0.11–0.18 (m, 2H), 0.40–0.53 (m, 2H), 0.82–0.95 (m, 1H), 1.85–1.96 (m, 2H), 4.38 (t, 1H, J = 6.3 Hz), 5.58 (d, 1H, J = 14.1 Hz), 5.62 (d, 1H, J = 14.1 Hz), 8.75 (d, 2H, J = 2.4 Hz), 8.87 (t, 1H, J = 2.4 Hz); ^{13}C NMR (75 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 4.88, 5.02, 7.24, 35.8, 54.3, 66.3, 119.2, 129.4, 140.9, 149.5, 170.2; HRMS electrospray ionization (ESI) calculated for $C_{13}H_{15}N_3O_6$: ($[M + H]^+$) m/z 310.1034, found: m/z 310.1032

(S)-3,5-Dinitrobenzyl 2-Amino-3-Cyclopentylpropanoate Hydrochloride (for the Introduction of Five-Membered-Ring Side Chain). The compound was obtained according to procedure A described above. ^{1}H NMR (300 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 1.06–1.22 (m, 2H), 1.41–1.63 (m, 4H), 1.75–1.89 (m, 2H), 1.96–2.08 (m, 1H), 4.22 (t, 2H, J = 7.5 Hz), 5.56 (d, 1H, J = 14.1 Hz), 5.61 (d, 1H, J = 14.1 Hz), 8.73 (d, 2H, J = 2.1 Hz), 8.89 (t, 1H, J = 2.1 Hz); ^{13}C NMR (75 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 25.4, 25.5, 32.8, 33.1, 36.8, 37.8, 53.5, 66.2, 119.2, 129.3, 140.9, 149.5, 170.9; HRMS (ESI) calculated for $C_{15}H_{19}N_3O_6$: ($[M + H]^+$) m/z 338.1347, found: m/z 338.1344

(S)-3,5-Dinitrobenzyl 2-Amino-3-Cyclohexylpropanoate Hydrochloride (for the Introduction of Six-Membered-Ring Side Chain). The compound was obtained according to procedure A described above. ^{1}H NMR (300 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 0.85–1.03 (m, 2H), 1.05–1.32 (m, 3H), 1.52–1.69 (m, 4H), 1.72–1.85 (m, 2H), 1.85–2.03 (m, 2H), 4.30 (t, 1H, J = 7.2 Hz), 5.57 (d, 1H, J = 13.8 Hz), 5.62 (d, 1H, J = 13.8 Hz), 8.77 (d, 2H, J = 2.1 Hz), 8.90 (t, 1H, J = 2.1 Hz); ^{13}C NMR (75 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 26.5, 26.6, 26.9, 33.3, 33.6, 34.3, 51.7, 66.2, 119.2, 129.4, 141.0, 149.5, 170.8; RMS (ESI) calculated for $C_{16}H_{21}N_3O_6$: ($[M + H]^+$) m/z 352.1503, found: m/z 352.1497

(S)-3,5-Dinitrobenzyl 2-Aminobutanoate Hydrochloride (for the Introduction of Two-Carbon Side Chain). The compound was obtained according to procedure A described above. ^{1}H NMR (300 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 1.09 (t, 3H, J = 7.2 Hz), 2.05–2.22 (m, 2H), 4.35 (t, 1H, J = 6.3 Hz), 5.62 (d, 1H, J = 13.8 Hz), 5.67 (d, 1H, J = 13.8 Hz), 8.78 (d, 2H, J = 1.8 Hz), 8.94 (t, 1H, J = 1.8 Hz); ^{13}C NMR (75 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 9.62, 24.3, 54.8, 66.4, 119.2, 129.4, 140.8, 149.4, 170.0; HRMS (ESI) calculated for $C_{11}H_{13}N_3O_6$: ($[M + H]^+$) m/z 284.0877, found: m/z 284.0877

(S)-3,5-Dinitrobenzyl 2-Aminopentanoate Hydrochloride (for the Introduction of Three-Carbon Side Chain). The compound was obtained according to procedure A described above. ^{1}H NMR (300 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 0.97 (t, 3H, J = 7.2 Hz), 1.39–1.69 (m, 2H), 1.98–2.16 (m, 2H), 4.38 (t, 1H, J = 6.3 Hz), 5.62 (d, 1H, J = 13.5 Hz), 5.67 (t, 1H, J = 13.5 Hz), 8.78 (d, 2H, J = 2.1 Hz), 8.93 (t, 1H, J = 2.1 Hz); ^{13}C NMR (75 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 13.8, 18.8, 32.9, 53.6, 66.4, 119.1, 129.3, 140.9, 149.5, 170.1; HRMS (ESI) calculated for $C_{12}H_{15}N_3O_6$: ($[M + H]^+$) m/z 298.1034, found: m/z 298.1031

(S)-3,5-Dinitrobenzyl 2-Aminohexanoate Hydrochloride (for the Introduction of Four-Carbon Side Chain). The compound was obtained according to procedure A described above. ^{1}H NMR (300 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 0.88 (t, 3H, J = 7.2 Hz), 1.30–1.63 (m, 4H), 2.09–2.14 (m, 2H), 4.31 (t, 1H, J = 6.3 Hz), 5.63 (s, 2H), 8.81 (d, 2H, J = 2.4 Hz), 8.92 (t, 1H, J = 2.4 Hz); ^{13}C NMR (75 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 13.3, 22.8, 27.7, 30.7, 53.7, 66.2, 119.2, 129.5, 141.0, 148.6, 169.1; HRMS (ESI) calculated for $C_{13}H_{17}N_3O_6$: ($[M + H]^+$) m/z 312.1190, found: m/z 312.1189

(S)-3,5-Dinitrobenzyl 2-Aminoheptanoate Hydrochloride (for the Introduction of Five-Carbon Side Chain). The compound was obtained according to procedure A described above. ^{1}H NMR (300 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 0.84 (t, 3H, J = 6.6 Hz), 1.19–1.56 (m, 6H), 2.01–2.16 (m, 2H), 4.38 (t, 1H, J = 6.3 Hz), 5.61 (d, 1H, J = 13.2 Hz), 5.69 (d, 1H, J = 13.2 Hz), 8.78 (s, 2H), 8.94 (s, 1H); ^{13}C NMR (75 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 14.3, 23.0, 25.4, 29.3, 30.9, 32.0, 53.7, 66.4, 119.2, 129.4, 140.9, 149.4, 170.1; HRMS (ESI) calculated for $C_{16}H_{23}N_3O_6$: ($[M + H]^+$) m/z 326.1347, found: m/z 326.1342

(S)-3,5-Dinitrobenzyl 2-Aminooctanoate Hydrochloride (for the Introduction of Six-Carbon Side Chain). The compound was obtained according to procedure A described above. ^{1}H NMR (300 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 0.83 (t, 3H, J = 6.9 Hz), 1.13–1.63 (m, 8H), 2.02–2.15 (m, 2H), 4.38 (t, 1H, J = 6.3 Hz), 5.61 (d, 1H, J = 13.5 Hz), 5.70 (d, 1H, J = 13.5 Hz), 8.78 (d, 2H, J = 2.4 Hz), 8.94 (t, 1H, J = 2.4 Hz); ^{13}C NMR (75 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 14.3, 23.0, 25.4, 29.3, 30.9, 32.0,

53.7, 66.4, 119.2, 129.4, 140.9, 149.4, 170.1; HRMS (ESI) calculated for $C_{16}H_{23}N_3O_6$: $([M+H]^+)$ m/z 340.1503, found: m/z 340.1495

(S)-3,5-Dinitrobenzyl 2-Aminononanoate Hydrochloride (for the Introduction of Seven-Carbon Side Chain). The compound was obtained according to procedure A described above. ^{1}H NMR (300 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 0.80 (t, 3H, J = 6.9 Hz), 1.19–1.32 (m, 12H), 4.10 (t, 1H, J = 6.3 Hz), 5.48 (d, 1H, J = 12.9 Hz), 5.52 (d, 1H, J = 12.9 Hz), 8.71 (s, 2H), 8.89 (s, 1H); ^{13}C NMR (75 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 14.3, 23.2, 25.8, 29.7, 29.9, 32.4, 32.8, 54.2, 65.7, 119.1, 128.5, 129.2, 141.4, 149.5; HRMS (ESI) calculated for $C_{16}H_{23}N_3O_6$: $([M+H]^+)$ m/z 354.1660, found: m/z 354.1657

(S)-3,5-Dinitrobenzyl 2-Aminodecanoate Hydrochloride (for the Introduction of Eight-Carbon Side Chain). The compound was obtained according to procedure A described above. ^{1}H NMR (300 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 0.688 (t, 3H, J = 7.2 Hz), 0.95–1.33 (m, 12H), 1.90 (dt, 2H, J = 9.3, 6.0 Hz), 4.21 (t, 1H, J = 6.0 Hz), 5.39 (d, 1H, J = 13.2 Hz), 5.54 (d, 1H, J = 13.2 Hz), 8.65 (d, 2H, J = 1.8 Hz), 8.96 (t, 1H, J = 1.8 Hz); ^{13}C NMR (75 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 13.4, 22.0, 23.9, 28.1, 28.28, 28.35, 29.7, 31.1, 52.7, 65.7, 119.0, 128.8, 139.2, 148.4, 169.8; HRMS (ESI) calculated for $C_{17}H_{25}N_3O_6$: $([M+H]^+)$ m/z 368.1816, found: m/z 368.1812

General Procedure for the Synthesis of Amino Acids with Cyanomethyl Ester (CME) (*Procedure B*). *(S)-Cyanomethyl 2-Amino-3-(1-Methyl-1H-indol-3-yl) Propanoate Trifluoroacetate (for the Introduction of Me-Trp).* The following is a representative experimental procedure of Procedure B: To a solution of 2-amino-3-(1-methyl-1*H*-indol-3-yl)propanoic acid (50 mg, 0.23 mmol) and *N*-[2-(trimethylsilyl) ethoxycarbonyl-oxy]succinimide (79 mg, 0.31 mmol) in H_2O (0.4 mL) and THF (0.4 mL), $NaHCO_3$ (48 mg, 0.57 mmol) was added at room temperature. After stirring for 5 h, the reaction mixture was diluted by adding EtOAc. The mixture was washed by saturated aqueous $KHSO_4$ and extracted with EtOAc (x3). The combined organic extracts were washed with brine, dried ($MgSO_4$), and concentrated in vacuo. To a solution of this crude mixture in DMF (0.8 mL), chloroacetonitrile (25 mg, 0.021 mL, 0.33 mmol) and triethylamine (76 mg, 0.10 mL, 0.75 mmol) were added at room temperature. After stirring for 16 h, the reaction was quenched by adding saturated aqueous $KHSO_4$, and mixture was extracted with Et_2O (x2). The combined organic extracts were washed with water (x2), saturated $NaHCO_3$ (x2) and brine, dried ($MgSO_4$), and concentrated in vacuo. The residue was purified by PTLC (hexane/EtOAc = 2/1) to afford CME derivative (72 mg, 78%). To a solution of DNB derivative (20 mg, 0.051 mmol) in EtOAc (0.5 mL), CME derivative (72 mg, 78%). To a solution of CME derivative (13 mg, 0.032 mmol) in CH_2Cl_2 (1 mL) was added trifluoroacetic acid (2 mL) at 0 °C. After stirring for 10 h, starting material was disappeared. The reaction mixture was concentrated in vacuo (13 mg, quantitative). ^{1}H NMR (300 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 3.50 (dd, 1H, J = 6.6, 2.4 Hz), 3.79 (s, 3H), 4.60 (t, 1H, J = 6.6 Hz), 4.97 (d, 1H, J = 15.9 Hz), 5.57 (d, 1H, J = 15.9 Hz), 7.18 (ddd, 1H, J = 8.1, 7.2, 0.9 Hz), 7.24 (s, 1H), 7.30 (ddd, 1H, J = 8.1, 7.2, 0.9 Hz), 7.46 (d, 1H, J = 8.1 Hz), 7.62 (d, 1H, J = 8.1 Hz); ^{13}C NMR (75 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 25.7, 32.2, 50.2, 53.2, 104.7, 110.1, 114.7, 116.6 (q, 1C, J = 288 Hz), 118.2, 119.6, 122.2, 127.0, 129.5, 137.1, 161.9 (q, 1C, J = 35 Hz), 168.5; HRMS (ESI) calculated for $C_{14}H_{15}N_3O_2$: $([M+H]^+)$ m/z 258.1237, found: m/z 258.1227

(S)-Cyanomethyl 2-Amino-3-(1H-Indol-3-yl) Propanoate Trifluoroacetate (for the Introduction of Trp). The compound was obtained according to procedure B described above. ^{1}H NMR (300 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 2.78 (s, 1H), 3.54 (dd, 1H, J = 15.3, 6.6 Hz), 3.59 (dd, 1H, J = 15.3, 6.6 Hz), 4.63 (t, 1H, J = 6.6 Hz), 5.01 (d, 1H, J = 15.9 Hz), 5.07 (d, 1H, J = 15.9 Hz), 7.22 (ddd, 1H, J = 8.1, 6.9, 1.5 Hz), 7.31 (ddd, 1H, J = 8.1, 6.9, 0.9 Hz), 7.36 (s, 1H), 7.56 (dd, 1H, J = 6.9, 1.5 Hz), 7.65 (dd, 1H, J = 6.9, 1.5 Hz); ^{13}C NMR (75 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 25.6, 50.3, 53.2, 105.7, 112.1, 116.4 (q, 1C, J = 290 Hz), 114.4, 117.9, 119.7, 122.3, 125.4, 126.4, 136.3, 168.6; HRMS (ESI) calculated for $C_{13}H_{13}N_3O_2$: $([M+H]^+)$ m/z 244.1081, found: m/z 244.1080

(S)-Cyanomethyl 2-Amino-3-(4-Methoxyphenyl) Propanoate Trifluoroacetate (for the Introduction of Me-Tyr). The compound was obtained according to procedure B described above. ^{1}H NMR (300 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 3.16 (d, 1H, J = 6.6 Hz), 3.70 (s, 3H), 4.41 (t, 1H, J = 6.6 Hz), 4.90 (d, 1H, J = 15.9 Hz), 4.96 (d, 1H, J = 15.9 Hz), 6.86–6.89 (m, 2H), 7.10–7.14 (m, 2H); ^{13}C NMR (75 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 34.7, 50.3, 53.9, 55.3, 114.7, 116.4 (q, 1C, J = 290 Hz), 125.6, 130.8, 158.7, 162.7 (q, 1C, J = 35 Hz), 168.3; HRMS (ESI) calculated for $C_{12}H_{14}N_2O_3$: $([M+H]^+)$ m/z 235.1077, found: m/z 235.1076

(S)-Cyanomethyl 2-Amino-3-(4-Hydroxyphenyl) Propanoate Trifluoroacetate (for the Introduction of Tyr). The compound was obtained according to procedure B described above. ^{1}H NMR (300 MHz, D_2O, the minor rotamer were marked with an asterisk) δ 3.03–3.25 (m, 2H), 4.27–4.44 (m, 1H), 4.48* (d, 0.3H, J = 14.4 Hz), 4.54* (d, 0.3H, J = 14.4 Hz), 4.91 (d, 0.7H, J = 15.9 Hz), 4.97 (d, 0.7H, J = 15.9 Hz), 6.75–6.82 (m, 2H), 7.03–7.11 (m, 2H); ^{13}C NMR (75 MHz, D_2O) δ 34.6, 50.3, 53.9, 114.7, 115.89, 115.97, 116.2 (q, 1C, J = 290 Hz), 124.9, 130.77, 130.82, 155.3, 162.7 (q, 1C, J = 35 Hz), 168.3; HRMS (ESI) calculated for $C_{12}H_{14}N_2O_3$: $([M+H]^+)$ m/z 221.0921, found: m/z 221.0933

(S)-Cyanomethyl 2-Amino-3-(Benzo[b]thiophen-3-yl) Propanoate Trifluoroacetate (for the Introduction of Benzothiophene Side Chain). The compound was obtained according to procedure B described above. ^{1}H NMR (300 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 3.60 (dd, 1H, J = 14.4, 7.2 Hz), 3.67 (dd, 1H, J = 14.4, 7.2 Hz), 4.67 (t, 1H, J = 7.2 Hz), 4.96 (d, 1H, J = 15.9 Hz), 5.02 (d, 1H, J = 15.9 Hz), 7.45–7.56 (m, 2H), 7.57 (s, 1H), 7.85 (dd, 1H, J = 6.6, 1.2 Hz), 8.01 (dd, 1H, J = 6.9, 1.2 Hz); ^{13}C NMR (75 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 28.4, 50.4, 52.4, 114.5, 116.4 (q, 1C, J = 290 Hz), 121.1, 123.3, 124.7, 125.0, 126.8, 127.6, 137.7, 140.3, 162.7 (q, 1C, J = 35 Hz), 168.2; HRMS (ESI) calculated for $C_{13}H_{12}N_2O_2S$: $([M+H]^+)$ m/z 261.0692, found: m/z 261.0684

(S)-Cyanomethyl 2-Amino-3-(4-Aminophenyl) Propanoate Trifluoroacetate (for the Introduction of Aniline Side Chain). The compound was obtained according to procedure B described above. ^{1}H NMR (300 MHz, DMSO-d_6/D_2O, 10/1, vol/vol, the minor rotamer was marked with an asterisk) δ 3.14–3.34 (m, 2H), 4.35–4.48 (m, 1H), 4.53* (d, 0.4H, J = 15.3 Hz), 4.60* (d, 0.4H, J = 15.3 Hz), 5.07 (s, 1.2H), 7.33–7.51(m, 4H); ^{13}C NMR (75 MHz, DMSO-d_6/D_2O, 10/1, vol/vol) δ 34.9, 50.3, 53.4, 114.6, 116.4 (q, 1C, J = 288 Hz), 118.3, 131.0, 131.2, 134.7, 168.1; HRMS (ESI) calculated for $C_{13}H_{12}N_2O_2S$: $([M+H]^+)$ m/z 220.1081, found: m/z 220.1081

(S)-Cyanomethyl 2-Amino-3-(Biphenyl-4-yl) Propanoate Trifluoroacetate (for the Introduction of Biphenyl). The compound was obtained according to procedure B described above. ^{1}H NMR (300 MHz, acetone-d_6/D_2O, 10/1, vol/vol, the minor rotamer was marked with an asterisk) δ 3.44–3.65 (m, 2H), 4.76* (br, 0.25H), 5.02–5.15

(m, 2H), 5.35 (dd, 0.75H, $J = 9.9, 4.5$ Hz), 7.33–7.41 (m, 1H), 7.44–7.52 (m, 4H), 7.61–7.70 (m, 4H); ^{13}C NMR (75 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 36.9, 50.8, 62.7, 115.3, 127.7, 127.99, 128.05, 128.2, 128.3, 129.71, 129.75, 129.77, 131.6, 135.8, 140.9, 141.2, 169.2; HRMS (ESI) calculated for $C_{17}H_{16}N_2O_2$: ($[M + H]^+$) m/z 281.1285, found: m/z 281.1285

(S)-Cyanomethyl 2-Amino-3-(Naphthalen-1-yl)Propanoate Trifluoroacetate (for the Introduction of 1-Naphtyl Side Chain). The compound was obtained according to procedure B described above. ^{1}H NMR (300 MHz, acetone-d_6/D_2O, 10/1, vol/vol, the minor rotamer was marked with an asterisk) δ 3.70–3.87 (m, 1H), 3.96–4.11 (m, 1H), 4.70 (t, 0.7H, $J = 7.5$ Hz), 4.89 (d, 0.7H, $J = 16.2$ Hz), 4.95 (d, 0.7H, $J = 16.2$ Hz), 5.05* (d, 0.3H, $J = 16.5$ Hz), 5.11* (d, 0.3H, $J = 16.5$ Hz), 5.22* (dd, 0.3H, $J = 10.2, 3.9$ Hz), 7.43–7.64 (m, 4H), 7.89 (d, 1H, $J = 8.1$ Hz), 7.96 (dd, 1H, $J = 8.1, 1.2$ Hz), 8.33 (dd, 1H, $J = 7.5, 7.2$ Hz); ^{13}C NMR (75 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 34.6, 50.6, 54.3, 115.1, 124.1, 126.4, 126.9, 127.6, 129.4, 129.4, 129.9, 131.5, 132.7, 134.9, 169.3; HRMS (ESI) calculated for $C_{15}H_{14}N_2O_2$: ($[M + H]^+$) m/z 255.1128, found: m/z 255.1119

(S)-Cyanomethyl 2-Amino-3-(Naphthalen-6-yl) Propanoate Trifluoroacetate (for the Introduction of 2-Naphtyl Side Chain). The compound was obtained according to procedure B described above. ^{1}H NMR (300 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 3.43 (dd, 1H, $J = 15.3, 6.9$ Hz), 3.48 (dd, 1H, $J = 15.3, 6.9$ Hz), 4.67 (t, 1H, $J = 6.9$ Hz), 4.94 (d, 1H, $J = 16.2$ Hz), 5.03 (d, 1H, $J = 16.2$ Hz), 7.37 (dd, 1H, $J = 8.7, 1.8$ Hz), 7.46–7.54 (m, 2H), 7.74 (s, 1H), 7.78–7.92 (m, 3H); ^{13}C NMR (75 MHz, acetone-d_6/D_2O, 10/1, vol/vol) δ 35.8, 50.2, 53.8, 114.6, 116.5 (q, 1C, $J = 290$ Hz), 126.6, 126.8, 126.9, 127.75, 127.78, 128.5, 129.1, 130.9, 132.7, 133.2, 162.5 (q, 1C, $J = 35$ Hz), 168.2; HRMS (ESI) calculated for $C_{15}H_{14}N_2O_2$: ($[M + H]^+$) m/z 255.1128, found: m/z 255.1117

(S)-Cyanomethyl 2-Amino-3-Phenylpropanoate Hydrochloride (for the Introduction of Phe). The compound was obtained according to procedure B described above and also reported in ref. 1. ^{1}H NMR (300 MHz, DMSO-d_6, the minor rotamers was marked with asterisks) δ 3.08–3.35 (m, 2H), 4.28–4.44 (m, 1H), 4.45* (d, 0.2H, $J = 15.6$ Hz), 4.54* (d, 0.2H, $J = 15.6$ Hz), 4.76** (d, 0.2H, $J = 14.7$ Hz), 4.82** (d, 0.2H, $J = 14.7$ Hz), 5.02 (d, 0.6H, $J = 16.5$ Hz), 7.20–7.48 (m, 5H), 8.88 (s, 3H); ^{13}C NMR (75 MHz, DMSO-d_6) δ 35.6, 50.0, 53.0, 115.1, 127.3, 128.5, 128.6, 129.4, 129.5, 168.1; HRMS (ESI) calculated for $C_{11}H_{12}N_2O_2$: ($[M + H]^+$) m/z 284.0877, found: m/z 284.0877

1. Saito H, Kourouklis D, Suga H (2001) An in vitro evolved precursor tRNA with aminoacylation activity. *EMBO J* 20:1797–1806.

Table S1. Measured ΔG_{app} values and standard deviations (three or four measurements) in kcal/mol for all sequences and amino acids in the study

H segments with G flanks (helix length 19)

ID	*Amino acid*	*Sequence of test helix H3*	ΔG^{meas}_{app}	*SD*
	(*S*)-2-aminobutanoic acid			
1		GGPG*AAALAAAAAAAAALAAAAGPGG	0.72	0.05
2		GGPGAA*ALAAAAAAAAALAAAAGPGG	0.36	0.04
3		GGPGAAAAL*AAAAAAAALAAAAGPGG	0.52	0.03
4		GGPGAAAALAAAA*AAAALAAAAGPGG	0.55	0.12
5		GGPGAAAALAAAAAAAA*LAAAAGPGG	0.49	0.03
6		GGPGAAAALAAAAAAAAALA*AAGPGG	0.58	0.04
7		GGPGAAAALAAAAAAAAALAAA*GPGG	0.53	0.11
	(*S*)-2-aminopentanoic acid (L-norvaline)			
8		GGPG*AAALAAAAAAAAALAAAAGPGG	0.36	0.05
9		GGPGAA*ALAAAAAAAAALAAAAGPGG	0.31	0.03
10		GGPGAAAAL*AAAAAAAALAAAAGPGG	0.27	0.07
11		GGPGAAAALAAAA*AAAALAAAAGPGG	0.28	0.03
12		GGPGAAAALAAAAAAAA*LAAAAGPGG	0.36	0.03
13		GGPGAAAALAAAAAAAAALA*AAGPGG	0.36	0.03
14		GGPGAAAALAAAAAAAAALAAA*GPGG	0.41	0.06
	(*S*)-2-aminohexanoic acid (L-norleucine)			
15		GGPG*AAALAAAAAAAAALAAAAGPGG	0.12	0.01
16		GGPGAA*ALAAAAAAAAALAAAAGPGG	0.05	0.05
17		GGPGAAAAL*AAAAAAAALAAAAGPGG	−0.02	0.03
18		GGPGAAAALAAAA*AAAALAAAAGPGG	0.11	0.04
19		GGPGAAAALAAAAAAAA*LAAAAGPGG	0.30	0.06
20		GGPGAAAALAAAAAAAAALA*AAGPGG	0.21	0.07
21		GGPGAAAALAAAAAAAAALAAA*GPGG	0.23	0.03
	(*S*)-2-aminoheptanoic acid			
22		GGPG*AAALAAAAAAAAALAAAAGPGG	−0.02	0.09
23		GGPGAA*ALAAAAAAAAALAAAAGPGG	−0.28	0.07
24		GGPGAAAAL*AAAAAAAALAAAAGPGG	−0.34	0.04
25		GGPGAAAALAAAA*AAAALAAAAGPGG	−0.28	0.10
26		GGPGAAAALAAAAAAAA*LAAAAGPGG	−0.07	0.05
27		GGPGAAAALAAAAAAAAALA*AAGPGG	0.06	0.05
28		GGPGAAAALAAAAAAAAALAAA*GPGG	0.03	0.09

H segments with G flanks (helix length 19)

ID	*Amino acid*	*Sequence of test helix H3*	ΔG^{meas}_{app}	*SD*
	(*S*)-2-aminooctanoic acid			
29		GGPG*AAALAAAAAAAAALAAAAGPGG	−0.08	0.06
30		GGPGAA*ALAAAAAAAAALAAAAGPGG	−0.48	0.11
31		GGPGAAAAL*AAAAAAAALAAAAGPGG	−0.57	0.05
32		GGPGAAAALAAAA*AAAALAAAAGPGG	−0.62	0.07
33		GGPGAAAALAAAAAAAA*LAAAAGPGG	−0.26	0.07
34		GGPGAAAALAAAAAAAAALA*AAGPGG	0.23	0.05
35		GGPGAAAALAAAAAAAAALAAA*GPGG	0.16	0.06
	(*S*)-2-aminononanoic acid			
36		GGPG*AAALAAAAAAAAALAAAAGPGG	−0.23	0.05
37		GGPGAA*ALAAAAAAAAALAAAAGPGG	−0.47	0.11
38		GGPGAAAAL*AAAAAAAALAAAAGPGG	−0.64	0.06
39		GGPGAAAALAAAA*AAAALAAAAGPGG	−0.79	0.06
40		GGPGAAAALAAAAAAAA*LAAAAGPGG	−0.48	0.07
41		GGPGAAAALAAAAAAAAALA*AAGPGG	−0.36	0.00
42		GGPGAAAALAAAAAAAAALAAA*GPGG	−0.29	0.07
	(*S*)-2-aminodecanoic acid			
43		GGPG*AAALAAAAAAAAALAAAAGPGG	−0.23	0.08
44		GGPGAA*ALAAAAAAAAALAAAAGPGG	−0.71	0.17
45		GGPGAAAAL*AAAAAAAALAAAAGPGG	−0.69	0.08
46		GGPGAAAALAAAA*AAAALAAAAGPGG	−0.96	0.10
47		GGPGAAAALAAAAAAAA*LAAAAGPGG	−0.70	0.06
48		GGPGAAAALAAAAAAAAALA*AAGPGG	−0.55	0.06
49		GGPGAAAALAAAAAAAAALAAA*GPGG	−0.30	0.06
	(*S*)-2-amino-3-cyclopropylpropanoic acid			
50		GGPG*AAALAAAAAAAAALAAAAGPGG	0.43	0.02
51		GGPGAA*ALAAAAAAAAALAAAAGPGG	0.4	0.04
52		GGPGAAAAL*AAAAAAAALAAAAGPGG	0.32	0.01
53		GGPGAAAALAAAA*AAAALAAAAGPGG	0.33	0.03
54		GGPGAAAALAAAAAAAA*LAAAAGPGG	0.57	0.09
55		GGPGAAAALAAAAAAAAALA*AAGPGG	0.43	0.07
56		GGPGAAAALAAAAAAAAALAAA*GPGG	0.32	0.04
	(*S*)-2-amino-3-cyclopentylpropanoic acid			
57		GGPG*AAALAAAAAAAAALAAAAGPGG	0.10	0.08
58		GGPGAA*ALAAAAAAAAALAAAAGPGG	0.00	0.05
59		GGPGAAAAL*AAAAAAAALAAAAGPGG	−0.21	0.12
60		GGPGAAAALAAAA*AAAALAAAAGPGG	−0.06	0.08
61		GGPGAAAALAAAAAAAA*LAAAAGPGG	0.01	0.05
62		GGPGAAAALAAAAAAAAALA*AAGPGG	0.13	0.02
63		GGPGAAAALAAAAAAAAALAAA*GPGG	−0.09	0.02
	(*S*)-2-amino-3-cyclohexylpropanoic acid			
64		GGPG*AAALAAAAAAAAALAAAAGPGG	0.35	0.04
65		GGPGAA*ALAAAAAAAAALAAAAGPGG	0.03	0.01
66		GGPGAAAAL*AAAAAAAALAAAAGPGG	0.03	0.04
67		GGPGAAAALAAAA*AAAALAAAAGPGG	−0.09	0.09
68		GGPGAAAALAAAAAAAA*LAAAAGPGG	0.17	0.05
69		GGPGAAAALAAAAAAAAALA*AAGPGG	0.15	0.05
70		GGPGAAAALAAAAAAAAALAAA*GPGG	0.11	0.08
	L-phenylalanine			
71		GGPG*AAALAAAAAAAAALAAAAGPGG	0.30	0.03
72		GGPGAA*ALAAAAAAAAALAAAAGPGG	0.32	0.06
73		GGPGAAAAL*AAAAAAAALAAAAGPGG	0.14	0.06
74		GGPGAAAALAAAA*AAAALAAAAGPGG	0.29	0.05
75		GGPGAAAALAAAAAAAA*LAAAAGPGG	0.35	0.07
76		GGPGAAAALAAAAAAAAALA*AAGPGG	0.29	0.02
77		GGPGAAAALAAAAAAAAALAAA*GPGG	0.19	0.02
	O-methyl-L-tyrosine			
78		GGPG*AAALAAAAAAAAALAAAAGPGG	0.46	0.08
79		GGPGAA*ALAAAAAAAAALAAAAGPGG	0.40	0.05
80		GGPGAAAAL*AAAAAAAALAAAAGPGG	0.28	0.03
81		GGPGAAAALAAAA*AAAALAAAAGPGG	0.28	0.06
82		GGPGAAAALAAAAAAAA*LAAAAGPGG	0.44	0.04
83		GGPGAAAALAAAAAAAAALA*AAGPGG	0.29	0.03
84		GGPGAAAALAAAAAAAAALAAA*GPGG	0.22	0.03

H segments with G flanks (helix length 19)

ID	*Amino acid*	*Sequence of test helix H3*	ΔG^{meas}_{app}	*SD*
	(*S*)-2-amino-3-(4-aminophenyl)propanoic acid			
85		GGPG*AAALALAAAAALALAAAAGPGG	−0.71	0.01
86		GGPGAA*ALALAAAAALALAAAAGPGG	−0.41	0.04
87		GGPGAAAAL*LAAAAALALAAAAGPGG	−0.06	0.09
88		GGPGAAAALALAA*AALALAAAAGPGG	0.15	0.07
89		GGPGAAAALALAAAAAL*LAAAAGPGG	−0.39	0.01
90		GGPGAAAALALAAAAALALA*AAGPGG	−0.94	0.13
91		GGPGAAAALALAAAAALALAAA*GPGG	−0.76	0.08
	L-tyrosine			
92		GGPG*AAALAAAAAAAAALAAAAGPGG	0.60	0.12
93		GGPGAA*ALAAAAAAAAALAAAAGPGG	0.78	0.07
94		GGPGAAAAL*AAAAAAAALAAAAGPGG	0.88	0.11
95		GGPGAAAALAAAA*AAAALAAAAGPGG	1.04	0.11
96		GGPGAAAALAAAAAAAA*LAAAAGPGG	0.90	0.03
97		GGPGAAAALAAAAAAAAALA*AAGPGG	0.43	0.05
98		GGPGAAAALAAAAAAAAALAAA*GPGG	0.77	0.05
	L-tyrosine			
99		GGPG*AAALALAAAAALALAAAAGPGG	−0.82	0.12
100		GGPGAA*ALALAAAAALALAAAAGPGG	−0.62	0.01
101		GGPGAAAAL*LAAAAALALAAAAGPGG	−0.49	0.03
102		GGPGAAAALALAA*AALALAAAAGPGG	0.18	0.02
103		GGPGAAAALALAAAAAL*LAAAAGPGG	−0.62	0.08
104		GGPGAAAALALAAAAALALA*AAGPGG	−1.11	0.09
105		GGPGAAAALALAAAAALALAAA*GPGG	−0.90	0.07
	L-tryptophan			
106		GGPG*AAALAAAAAAAAALAAAAGPGG	0.07	0.04
107		GGPGAA*ALAAAAAAAAALAAAAGPGG	0.23	0.05
108		GGPGAAAAL*AAAAAAAALAAAAGPGG	0.66	0.05
109		GGPGAAAALAAAA*AAAALAAAAGPGG	0.77	0.05
110		GGPGAAAALAAAAAAAA*LAAAAGPGG	0.66	0.08
111		GGPGAAAALAAAAAAAAALA*AAGPGG	0.25	0.09
112		GGPGAAAALAAAAAAAAALAAA*GPGG	0.43	0.12
	L-tryptophan			
113		GGPG*AAALALAAAAALALAAAAGPGG	−1.41	0.17
114		GGPGAA*ALALAAAAALALAAAAGPGG	−1.04	0.08
115		GGPGAAAAL*LAAAAALALAAAAGPGG	−0.69	0.07
116		GGPGAAAALALAA*AALALAAAAGPGG	−0.37	0.09
117		GGPGAAAALALAAAAAL*LAAAAGPGG	−0.92	0.03
118		GGPGAAAALALAAAAALALA*AAGPGG	−1.47	0.18
119		GGPGAAAALALAAAAALALAAA*GPGG	−1.19	0.07
	(*S*)-2-amino-3-(1-methyl-1*H*-indol-3-yl)propanoic acid			
120		GGPG*AAALAAAAAAAAALAAAAGPGG	0.06	0.06
121		GGPGAA*ALAAAAAAAAALAAAAGPGG	0.15	0.03
122		GGPGAAAAL*AAAAAAAALAAAAGPGG	0.17	0.10
123		GGPGAAAALAAAA*AAAALAAAAGPGG	0.15	0.09
124		GGPGAAAALAAAAAAAA*LAAAAGPGG	0.41	0.09
125		GGPGAAAALAAAAAAAAALA*AAGPGG	0.09	0.02
126		GGPGAAAALAAAAAAAAALAAA*GPGG	0.17	0.08
	(*S*)-2-amino-3-(benzo[*b*]thiophen-3-yl)propaonoic acid			
127		GGPG*AAALAAAAAAAAALAAAAGPGG	−0.05	0.09
128		GGPGAA*ALAAAAAAAAALAAAAGPGG	−0.12	0.03
129		GGPGAAAAL*AAAAAAAALAAAAGPGG	−0.18	0.01
130		GGPGAAAALAAAA*AAAALAAAAGPGG	−0.20	0.01
131		GGPGAAAALAAAAAAAA*LAAAAGPGG	0.09	0.11
132		GGPGAAAALAAAAAAAAALA*AAGPGG	−0.02	0.03
133		GGPGAAAALAAAAAAAAALAAA*GPGG	−0.19	0.02
	(*S*)-2-amino-3-(naphthalen-1-yl)propanoic acid			
134		GGPG*AAALAAAAAAAAALAAAAGPGG	−0.11	0.00
135		GGPGAA*ALAAAAAAAAALAAAAGPGG	−0.14	0.02
136		GGPGAAAAL*AAAAAAAALAAAAGPGG	−0.20	0.04
137		GGPGAAAALAAAA*AAAALAAAAGPGG	−0.10	0.05
138		GGPGAAAALAAAAAAAA*LAAAAGPGG	0.03	0.01
139		GGPGAAAALAAAAAAAAALA*AAGPGG	−0.11	0.02

H segments with G flanks (helix length 19)

ID	*Amino acid*	*Sequence of test helix H3*	ΔG^{meas}_{app}	*SD*
140		GGPGAAAALAAAAAAAAALAAA*GPGG	−0.06	0.00
	(*S*)-2-amino-3-(naphthalen-6-yl) propanoic acid			
141		GGPG*AAALAAAAAAAAALAAAAGPGG	−0.03	0.08
142		GGPGAA*ALAAAAAAAAALAAAAGPGG	−0.04	0.04
143		GGPGAAAAL*AAAAAAAALAAAAGPGG	−0.15	0.00
144		GGPGAAAALAAAA*AAAALAAAAGPGG	−0.24	0.04
145		GGPGAAAALAAAAAAAA*LAAAAGPGG	0.01	0.05
146		GGPGAAAALAAAAAAAALA*AAGPGG	−0.07	0.03
147		GGPGAAAALAAAAAAAALAAA*GPGG	−0.03	0.06
	(*S*)-2-amino-3-(biphenyl-4-yl) propanoic acid			
148		GGPG*AAALAAAAAAAAALAAAAGPGG	−0.06	0.02
149		GGPGAA*ALAAAAAAAAALAAAAGPGG	−0.15	0.05
150		GGPGAAAAL*AAAAAAAALAAAAGPGG	−0.43	0.02
151		GGPGAAAALAAAA*AAAALAAAAGPGG	−0.37	0.07
152		GGPGAAAALAAAAAAAA*LAAAAGPGG	−0.07	0.02
153		GGPGAAAALAAAAAAAALA*AAGPGG	−0.07	0.00
154		GGPGAAAALAAAAAAAALAAA*GPGG	−0.21	0.07
	L-alanine			
155		GGPGAAAALAAAAAAAALAAAAGPGG	0.66	0.06
	L-valine			
156		GGPGAAAALAAAAVAAAALAAAAGPGG	0.18	0.03
	L-leucine			
157		GGPGAAAALAAAALAAAALAAAAGPGG	0.10	0.02
	L-isoleucine			
158		GGPGAAAALAAAAIAAAALAAAAGPGG	−0.07	0.05
	amino acid	*sequence of test helix H4*	ΔG^{meas}_{app}	*SD*
	(*S*)-2-aminooctanoic acid			
159		GGPG*AAAAAAAAAAAAAAAAAAGPGG	0.52	0.09
160		GGPGAA*AAAAAAAAAAAAAAAAGPGG	0.49	0.05
161		GGPGAAAAA*AAAAAAAAAAAAAGPGG	0.34	0.08
162		GGPGAAAAAAAAA*AAAAAAAAAGPGG	0.29	0.05
163		GGPGAAAAAAAAAAAA*AAAAAGPGG	0.18	0.03
164		GGPGAAAAAAAAAAAAAAAA*AAGPGG	0.36	0.10
165		GGPGAAAAAAAAAAAAAAAAAA*GPGG	0.29	0.10
	(*S*)-2-aminodecanoic acid			
166		GGPG*AAAAAAAAAAAAAAAAAAGPGG	0.16	0.07
167		GGPGAA*AAAAAAAAAAAAAAAAGPGG	0.31	0.05
168		GGPGAAAAA*AAAAAAAAAAAAAGPGG	0.19	0.05
169		GGPGAAAAAAAAA*AAAAAAAAAGPGG	0.04	0.04
170		GGPGAAAAAAAAAAAA*AAAAAGPGG	−0.03	0.11
171		GGPGAAAAAAAAAAAAAAAA*AAGPGG	−0.02	0.15
172		GGPGAAAAAAAAAAAAAAAAAA*GPGG	0.12	0.08
	L-alanine			
173		GGPGAAAAAAAAAAAAAAAAAAGPGG	1.20	0.10

Apolar surface area determines the efficiency of translocon-mediated membrane-protein integration into the endoplasmic reticulum

Karin Öjemalm[a], Takashi Higuchi[b], Yang Jiang[c], Ülo Langel[c], IngMarie Nilsson[a], Stephen H. White[d], Hiroaki Suga[b,1], and Gunnar von Heijne[a,e,1]

[a]Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden; [b]Department of Chemistry, Graduate School of Science, The University of Tokyo, 113-0033 Tokyo, Japan; [c]Department of Neurochemistry, Stockholm University, SE-10691 Stockholm, Sweden; [d]Department of Physiology and Biophysics and the Center for Biomembrane Systems, University of California, Irvine, CA 92697-4560; and [e]Science for Life Laboratory Stockholm University, SE-171 77 Solna, Sweden

AUTHOR SUMMARY

In eukaryotic cells, membrane proteins destined for the plasma membrane and many intracellular compartments are synthesized by endoplasmic reticulum (ER)-bound ribosomes—the cell's protein builders. Proteins are then integrated into the membrane of the ER in a process mediated by the Sec61 translocon complex. During the membrane-integration step, α-helical hydrophobic segments in the translocating nascent polypeptide chain exit the Sec61 translocon and become embedded in the surrounding lipid bilayer (1). Translocon-mediated integration of transmembrane α-helices into the ER membrane sets the stage for all subsequent folding and oligomerization events and hence represents a critical step in the maturation of membrane proteins. To clarify the nature of the membrane-integration process, we have measured the efficiency of translocon-mediated insertion of model hydrophobic segments containing nonproteinogenic amino acids into the ER membrane. We find that an amino acid's contribution to the apparent free energy of membrane insertion (ΔG_{app}) is directly proportional to the accessible nonpolar surface area of its side chain, as expected for thermodynamic partitioning between aqueous and nonpolar phases.

In previous studies, we provided quantitative data on the propensities of the 20 natural amino acids to promote the integration of transmembrane helices into the ER membrane and showed that they depend both on hydrophobicity and on position within the helix (2). Although the partitioning of transmembrane helices between the Sec61 translocon and the lipid membrane bears strong similarities to partitioning of solutes between water and lipid membranes, translocon-to-bilayer partitioning may not be equivalent to water-to-bilayer partitioning. Insights into the differences between the two partitioning processes might be revealed if the physicochemical properties of the translocon could be probed chemically. However, given the somewhat idiosyncratic collection of proteinogenic amino acids found in nature, it has hitherto not been possible to vary side-chain chemistry in the systematic fashion required to unravel fully the physicochemical basis for translocon-mediated membrane partitioning.

To probe the membrane integration mechanism in greater detail, we have taken advantage of a suppressor tRNA-based technique to introduce nonproteinogenic aliphatic and aromatic amino acids into model hydrophobic segments and measure their contribution to ΔG_{app}. Briefly, an engineered stop codon that can be recognized by the charged suppressor tRNA serves to position the nonproteinogenic amino acid in the hydrophobic segment. The suppressor tRNA is charged with the desired nonproteinogenic amino acid and is then added to an in vitro translation system. The system is programmed with an mRNA that encodes a host protein containing the engineered hydrophobic segment

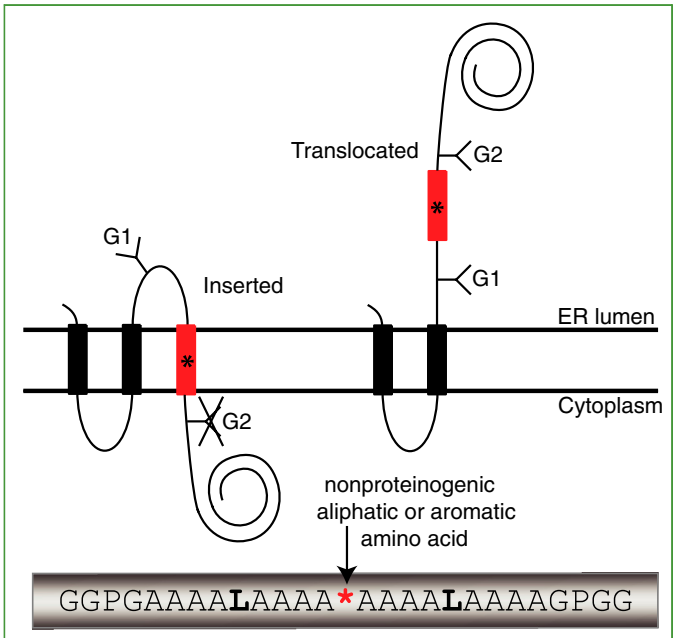

Fig. P1. Model hydrophobic segments (red) are introduced into a host membrane protein with two transmembrane segments (black), where they are placed between two engineered acceptor sites for N-linked glycosylation (G1, G2). A typical hydrophobic segment is shown, including a nonproteinogenic amino acid (*) and GGPG…GPGG flanking regions. Because the oligosaccharyl transferase enzyme is present only in the ER lumen, the G1 and G2 sites can only be glycosylated if they are translocated across the ER membrane to the lumenal side. The fractions of molecules with membrane-inserted (*Left*) and noninserted (*Right*) hydrophobic segment can therefore be determined by quantitation of radiolabeled singly and doubly glycosylated forms of the protein.

Author contributions: K.Ö., T.H., H.S., and G.v.H. designed research; K.Ö. and T.H. performed research; Y.J., U.L., and I.M.N. contributed new reagents/analytic tools; K.Ö., T.H., S.H.W., and G.v.H. analyzed data; and K.Ö., T.H., S.H.W., H.S., and G.v.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. E-mail: gunnar@dbb.su.se or hsuga@chem.s.u-tokyo.ac.jp.

See full research article on page E359 of www.pnas.org.

Cite this Author Summary as: PNAS 10.1073/pnas.1100120108.

(Fig. P1). To mimic the cellular environment, translation is carried out in the presence of ER-derived membrane vesicles, and the degree of glycosylation of two glycan acceptor sites that bracket the hydrophobic segment is used to calculate its degree of insertion into the vesicular membrane.

We find that an amino acid's contribution to ΔG_{app} is directly proportional to the accessible surface area of its side chain. We further find that long alkyl side chains tend to partition into the center of the membrane, as expected for transfer between bulk water and a lipid bilayer.

We have also studied the effect on ΔG_{app} of polar groups in aromatic side chains and find that methylation of hydroxyl or amine groups strongly reduces ΔG_{app} when the side chain is placed near the center of the membrane. This suggests that the hydrogen-bonding ability of polar aromatic side chains is critical for membrane insertion. An earlier study of the interactions of tryptophan analogs with and without hydrogen-bonding ability showed that the "aromaticity" of tryptophan was the dominant cause of its preferential partitioning into lipid bilayer interfaces (3). Our current results indicate that hydrogen bonding becomes important when polar aromatic residues are inserted in locations below the membrane-water interface.

Our study provides compelling evidence for the thermodynamic partitioning model of translocon-mediated integration of transmembrane helices into the ER membrane. In its simplest version, this model pictures the transmembrane helix as equilibrating between the translocon channel and the surrounding membrane. As judged from the available X-ray structures of prokaryotic homologs of the Sec61 translocon, the channel is quite narrow and lined by a mixture of polar and apolar amino acids, providing an environment that is less polar than aqueous buffer. Likewise, the ER membrane, which, like all biological membranes, has a high protein content, may offer an environment that is less apolar than a pure lipid bilayer. Indeed, the solvation parameters for partitioning of aliphatic and nonpolar aromatic side chains obtained here [approximately 10 and 7 cal/(mol $\cdot$ Å^2)] are 2.5 times smaller than those found in classical solute transfer experiments (4), which suggests that simple water-to-lipid partitioning measurements do not capture the full complexity of translocon-to-membrane partitioning.

1. van den Berg B, et al. (2004) X-ray structure of a protein-conducting channel. *Nature* 427:36–44.
2. Hessa T, et al. (2007) Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* 450:1026–1030.
3. Yau WM, Wimley WC, Gawrisch K, White SH (1998) The preference of tryptophan for membrane interfaces. *Biochemistry* 37:14713–14718.
4. Karplus PA (1997) Hydrophobicity regained. *Protein Sci* 6:1302–1307.